OECD GUIDELINE FOR THE TESTING OF CHEMICALS

DRAFT PROPOSAL FOR AN UPDATED TEST GUIDELINE 429

Skin Sensitisation: Local Lymph Node Assay

INTRODUCTION

1. OECD Guidelines for the Testing of Chemicals are periodically reviewed in light of scientific progress, changing regulatory needs, and animal welfare considerations. The first Test Guideline (TG) for the determination of skin sensitisation in the mouse, the Local Lymph Node Assay (LLNA; TG 429) was adopted in 2002 (1). The details of the validation of the LLNA and a review of the associated work have been published (2)(3)(4)(5)(6)(7)(8). The updated LLNA is based on the evaluation of experience and scientific data (9). This is the second TG to be promulgated for assessing skin sensitisation potential of chemicals in animals. The other TG (*i.e.* TG 406) utilises guinea pig tests, notably the guinea pig maximisation test and the Buehler test (10). The LLNA provides certain advantages over TG 406 (10) with regard to animal welfare. This updated LLNA TG includes a set of Performance Standards (PS) (Annex 1) that can be used to evaluate the validation status of new and/or modified test methods that are functionally and mechanistically similar to the LLNA, in accordance with the principles of Guidance Document No. 34 (11).

2. The LLNA studies the induction phase of skin sensitisation and provides quantitative data suitable for dose-response assessment. It should be noted that the mild/moderate sensitizers which are recommended as suitable positive control (PC) test substances for guinea pig test methods (*i.e.* TG 406) (10) are also appropriate for use with the LLNA (6)(8)(12). A reduced LLNA (rLLNA) protocol that uses fewer animals is also described in this TG (13)(14)(15). The rLLNA may be used for the hazard classification of skin sensitising test substances when dose-response information is not needed, provided there is adherence to all other LLNA protocol specifications, as described in this TG. The rLLNA should not be used for the hazard identification of skin sensitising test substances when dose-response information is needed.

DEFINITIONS

3. Definitions used are provided in Annex 2.

INITIAL CONSIDERATIONS AND LIMITATIONS

4. The LLNA provides an alternative method for identifying potential skin sensitising test substances. This does not necessarily imply that in all instances the LLNA should be used in place of guinea pig tests (*i.e.* TG 406) (10), but rather that the assay is of equal merit and may be employed as an alternative in which positive and negative results generally no longer require further confirmation. The testing laboratory should consider all available information on the test substance prior to conducting the study. Such information will include the identity and chemical structure of the test substance; its physicochemical properties; the

results of any other *in vitro* or *in vivo* toxicity tests on the test substance; and toxicological data on structurally related test substances.

5. The LLNA is an *in vivo* method and, as a consequence, will not eliminate the use of animals in the assessment of allergic contact sensitising activity. It has, however, the potential to reduce the number of animals required for this purpose. Moreover, the LLNA offers a substantial refinement of the way in which animals are used for allergic contact sensitisation testing. The LLNA is based upon consideration of immunological events stimulated by chemicals during the induction phase of sensitisation. Unlike guinea pig tests (*i.e.* TG 406) (10) the LLNA does not require that challenge-induced dermal hypersensitivity reactions be elicited. Furthermore, the LLNA does not require the use of an adjuvant, as is the case for the guinea pig maximisation test (10). Thus, the LLNA reduces animal distress. Despite the advantages of the LLNA over TG 406 (10), it should be recognised that there are certain limitations that may necessitate the use of TG 406 (10) (*e.g.* false negative findings in the LLNA with certain metals, false positive findings with certain skin irritants [such as some surfactant type materials] (16)(17), or solubility of the test material). In addition, test substance classes or materials containing functional groups shown to act as potential confounders (18) may necessitate the use of guinea pig tests (*i.e.* TG 406) (10). Other than such identified limitations, the LLNA should be applicable for testing any test substances unless there are properties associated with these materials that may interfere with the accuracy of the LLNA.

PRINCIPLE OF THE TEST

6. The basic principle underlying the LLNA is that sensitizers induce proliferation of lymphocytes in the lymph nodes draining the site of test substance application. This proliferation is proportional to the dose and to the potency of the applied allergen and provides a simple means of obtaining a quantitative measurement of sensitisation. Proliferation is measured by comparing the mean proliferation in each test group to the mean proliferation in the vehicle treated control (VC) group. The ratio of the mean proliferation in each treated group to that in the concurrent VC group, termed the Stimulation Index (SI), is determined, and should be ≥ 3 before further evaluation of the test substance as a potential skin sensitizer is warranted. The methods described here are based on the use of *in vivo* radioactive labelling to measure an increased number of proliferating cells in the draining auricular lymph nodes. However, other endpoints for assessment of the number of proliferating cells may be employed provided the PS requirements are fully met (Annex 2).

DESCRIPTION OF THE ASSAY

Selection of animal species

7. The mouse is the species of choice for this test. Young adult female mice of CBA/Ca or CBA/J strain, which are nulliparous and non-pregnant, are used. At the start of the study, animals should be between 8-12 weeks old, and the weight variation of the animals should be minimal and not exceed 20% of the mean weight. Alternatively, other strains and males may be used when sufficient data are generated to demonstrate that significant strain and/or gender-specific differences in the LLNA response do not exist.

Housing and feeding conditions

8. Mice should be group housed (19), unless adequate scientific rationale for housing mice individually is provided. The temperature of the experimental animal room should be 22ºC (± 3ºC). Although the relative humidity should be at least 30% and preferably not exceed 70%, other than during room cleaning, the aim should be 50-60%. Lighting should be artificial, the sequence being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water.

Preparation of animals

9. The animals are randomly selected, marked to permit individual identification (but not by any form of ear marking), and kept in their cages for at least five days prior to the start of dosing to allow for acclimatisation to the laboratory conditions. Prior to the start of treatment all animals are examined to ensure that they have no observable skin lesions.

Preparation of dosing solutions

10. Solid test substances should be dissolved or suspended in solvents/vehicles and diluted, if appropriate, prior to application to an ear of the mice. Liquid test substances may be applied neat or diluted prior to dosing. Insoluble materials, such as those generally seen in medical devices, should be subjected to an exaggerated extraction in an appropriate solvent to reveal all extractable constituents for testing prior to application to an ear of the mice. Test substances should be prepared daily unless stability data demonstrate the acceptability of storage.

Reliability check

11. Positive controls (PC) are used to demonstrate appropriate performance of the assay by responding with adequate and reproducible sensitivity to a sensitising test substance for which the magnitude of the response is well characterised. Inclusion of a concurrent PC is recommended because it demonstrates competency of the laboratory to successfully conduct each assay and allows for an assessment of intra- and inter-laboratory reproducibility and comparability. A PC for each study is also required by some regulatory authorities. Accordingly, the routine use of a concurrent PC is encouraged to avoid the need for additional animal testing to meet such requirements that might arise from the use of a periodic PC (see paragraph 12). The PC should produce a positive LLNA response at an exposure level expected to give an increase in the SI > 3 over the negative control (NC) group. The PC dose should be chosen such that the induction is reproducible but not excessive (*i.e.* SI > 20). Preferred PC test substances are 25% hexyl cinnamic aldehyde (Chemical Abstracts Service [CAS] No 101-86-0) in acetone: olive oil and 5% mercaptobenzothiazole (CAS No 149-30-4) in *N,N*-dimethylformamide (see Annex 1, Table 1). There may be circumstances in which, given adequate justification, other PC test substances, meeting the above criteria, may be used.

12. While inclusion of a concurrent PC group is recommended, there may be situations in which periodic testing (*i.e.* at intervals ≤6 months) of the PC test substance may be adequate for laboratories that conduct the LLNA regularly *(i.e.* conduct the LLNA at a

frequency of no less than once per month) and have an established historical PC database that demonstrates the laboratory's ability to obtain reproducible and accurate results with PCs. Adequate proficiency with the LLNA can be successfully demonstrated by generating consistent positive results with the PC in at least 10 independent tests conducted within a reasonable period of time (*i.e.* less than one year).

13. A concurrent PC group should always be included when there is a procedural change to the LLNA (*e.g.* change in trained personnel, change in test method materials and/or reagents, change in test method equipment, change in source of test animals), and such changes should be documented in laboratory reports. Consideration should be given to the impact of these changes on the adequacy of the previously established historical database in determining the necessity for establishing a new historical database to document consistency in the PC results.

14. Investigators should be aware that the decision to conduct a PC on a periodic basis instead of concurrently has ramifications on the adequacy and acceptability of negative study results generated without a concurrent PC during the interval between each periodic PC study. For example, if a false negative result is obtained in the periodic PC study, negative test substance results obtained in the interval between the last acceptable periodic PC study and the unacceptable periodic PC study may be questioned. Implications of these outcomes should be carefully considered when determining whether to include concurrent PCs or to only conduct periodic PCs. Consideration should also be given to using fewer animals in the concurrent PC group when this is scientifically justified and if the laboratory demonstrates, based on laboratory-specific historical data, that fewer mice can be used (9)

15. Although the PC test substance should be tested in the vehicle that is known to elicit a consistent response (*e.g.* acetone: olive oil), there may be certain regulatory situations in which testing in a non-standard vehicle (clinically/chemically relevant formulation) will also be necessary (20). If the concurrent PC test substance is tested in a different vehicle than the test substance, then a separate vehicle control for the concurrent PC should be included.

16. In instances where test substances of a specific chemical class or range of responses are being evaluated, benchmark test substances may also be useful to demonstrate that the test method is functioning properly for detecting the skin sensitisation potential of these types of test substances. Appropriate benchmark test substances should have the following properties:

- structural and functional similarity to the class of the test substance being tested;
- known physical/chemical characteristics;
- supporting data from the LLNA;
- supporting data from other animal models and/or from humans.

TEST PROCEDURE

Number of animals and dose levels

17. A minimum of four animals is used per dose group, with a minimum of three concentrations of the test substance, plus a concurrent negative control group treated only with the vehicle for the test substance, and a PC (concurrent or recent, based on laboratory policy in considering paragraphs 11-15). Except for absence of treatment with the test substance, animals in the control groups should be handled and treated in a manner identical to that of animals in the treatment groups.

18. Dose and vehicle selection should be based on the recommendations given in references (3) and (5). Consecutive doses are normally selected from an appropriate concentration series such as 100%, 50%, 25%, 10%, 5%, 2.5%, 1%, 0.5%, etc. Adequate scientific rationale should accompany the selection of the concentration series used. All existing toxicological information (*e.g.* acute toxicity and dermal irritation) and structural and physicochemical information on the test substance of interest (and/or structurally related test substances) should be considered where available, in selecting the three consecutive concentrations so that the highest concentration maximises exposure while avoiding systemic toxicity and/or excessive local skin irritation (3)(21). In the absence of such information, an initial pre-screen test may be necessary (see paragraphs 21-1).

19. The vehicle should not interfere with or bias the test result and should be selected on the basis of maximising the solubility in order to obtain the highest concentration achievable while producing a solution/suspension suitable for application of the test substance. Recommended vehicles are acetone: olive oil (4:1 v/v), *N*,*N*-dimethylformamide, methyl ethyl ketone, propylene glycol, and dimethyl sulphoxide (16) but others may be used if sufficient scientific rationale is provided. In certain situations it may be necessary to use a clinically relevant solvent or the commercial formulation in which the test substance is marketed as an additional control. Particular care should be taken to ensure that hydrophilic materials are incorporated into a vehicle system, which wets the skin and does not immediately run off by incorporation of appropriate solubilisers (*e.g.* 1% Pluronic® L92). Thus, wholly aqueous vehicles are to be avoided.

20. The processing of lymph nodes from individual mice allows for the assessment of inter-animal variability and a statistical comparison of the difference between test substance and vehicle control group measurements (see paragraph 34). In addition, evaluating the possibility of reducing the number of mice in the PC group is only feasible when individual animal data are collected (9). Further, some national regulatory authorities require the collection of individual animal data. Regular collection of individual animal data provides an animal welfare advantage by avoiding duplicate testing that would be necessary if the test substance results originally collected in one manner (*e.g.* via pooled animal data) were to be considered later by regulatory authorities with other requirements (*e.g.* individual animal data).

Pre-screen test

21. In the absence of information to determine the highest dose to be tested (see paragraph 18), a pre-screen test should be performed in order to define the appropriate dose level to test in the LLNA. The purpose of the pre-screen test is to provide guidance for selecting the maximum dose level to use in the main LLNA study, where information on the

concentration that induces systemic toxicity (see paragraph 1) and/or excessive local skin irritation (see paragraph 23) is not available. The maximum dose level tested should be 100% of the test substance for liquids or the maximum possible concentration for solids or suspensions, unless available information suggests that this concentration induces systemic toxicity and/or excessive local irritation after topical application in the mouse.

22. The pre-screen test is conducted under conditions identical to the main LLNA study, except there is no assessment of lymph node proliferation and fewer animals per dose group can be used. One or two animals per dose group are suggested. All mice will be observed daily for any clinical signs of systemic toxicity or local irritation at the application site. Body weights are recorded pre-test and prior to termination (Day 6). Both ears of each mouse are observed for erythema and scored using Table 1 (21). Ear thickness measurements are taken using a thickness gauge (*e.g.* digital micrometer or Peacock Dial thickness gauge) on Day 1 (pre-dose), Day 3 (approximately 48 hours after the first dose), and Day 6. Additionally, on Day 6, ear thickness could be determined by ear punch weight determinations. Excessive local skin irritation is indicated by an erythema score ≥3 and/or ear thickness of ≥25% on any day of measurement (22)(23). The highest dose selected for the main LLNA study will be the next lower dose in the pre-screen concentration series (see paragraph 18) that does not induce systemic toxicity and/or excessive local skin irritation.

Table 1. Erythema scores.

Observation	Score
No erythema	0
Very slight erythema (barely perceptible)	1
Well-defined erythema	2
Moderate to severe erythema	3
Severe erythema (beet redness) to eschar formation preventing grading of erythema	4

23. In addition to a 25% increase in ear thickness (22)(23), a statistically significant increase in ear thickness in the treated mice compared to control mice has also been used to identify irritants in the LLNA (24)(25)(26)(27)(28)(29)(30). However, while statistically significant increases can occur when ear thickness is less than 25% they have not been associated specifically with excessive irritation (26)(28)(29)(30).

24. The following clinical observations may indicate systemic toxicity (1)(32) when used as part of an integrated assessment and therefore may indicate the maximum dose level to use in the main LLNA: changes in nervous system function (*e.g.* pilo-erection, ataxia, tremors, and convulsions); changes in behaviour (*e.g.* aggressiveness, change in grooming activity, marked change in activity level); changes in respiratory patterns (*i.e.* changes in frequency and intensity of breathing such as dyspnea, gasping, and rales), and changes in food and water consumption. In addition, signs of lethargy and/or unresponsiveness and any clinical signs of more than slight or momentary pain and distress, or a >5% reduction in body weight from Day 1 to Day 6, and mortality should be considered in the evaluation.

Reduced LLNA

25. Use of an rLLNA protocol (13)(14)(15) instead of the traditional multi-dose LLNA has the potential to reduce the number of animals used in a test by omitting the middle and low dose groups. The reduction in number of dose groups is the only difference between the LLNA and the rLLNA test method protocols and for this reason the rLLNA does not provide dose-response information. Therefore, the rLLNA should not be used when dose-response information is needed. Like the multi-dose LLNA, the test substance concentration evaluated in the rLLNA should be the maximum concentration that does not induce overt systemic toxicity and/or excessive local skin irritation in the mouse (see paragraph 18).

Main study experimental schedule

26. The experimental schedule of the assay is as follows:

- *Day 1:*
  Individually identify and record the weight of each animal and any clinical observations. Apply 25 µL of the appropriate dilution of the test substance, the vehicle alone, or the concurrent PC (see paragraphs 11-15), to the dorsum of each ear.
- *Days 2 and 3:*
  Repeat the application procedure carried out on Day 1.
- *Days 4 and 5:*
  No treatment.
- *Day 6:*
  Record the weight of each animal. Inject 250 µL of sterile phosphate-buffered saline (PBS) containing 20 µCi (7.4×10^5 Bq) of tritiated (^{3}H)-methyl thymidine into all test and control mice via the tail vein. Alternatively, inject 250 µL sterile PBS containing 2 µCi (7.4×10^4 Bq) of ^{125}I-iododeoxyuridine and 10^{-5}M fluorodeoxyuridine into all mice via the tail vein. Five hours (5 h) later, humanely kill the animals. Excise the draining auricular lymph nodes from each mouse ear and process separately in PBS for each animal. Details and diagrams of the node identification and dissection can be found in reference (9). To further monitor the local skin response in the main study, additional parameters such as scoring of ear erythema or ear thickness measurements (obtained either by using a thickness gauge, or ear punch weight determinations at necropsy) may be included in the study protocol.

Preparation of cell suspensions

27. From each mouse, a single-cell suspension of lymph node cells (LNC) excised bilaterally is prepared by gentle mechanical disaggregation through 200 micron-mesh stainless steel gauze or another acceptable technique for generating a single-cell suspension. LNC are washed twice with an excess of PBS and the DNA is precipitated with 5% trichloroacetic acid (TCA) at 4^{o}C for 18h (3). Pellets are either re-suspended in 1 mL TCA and transferred to scintillation vials containing 10 mL of scintillation fluid for ^{3}H-counting, or transferred directly to gamma counting tubes for ^{125}I-counting.

Determination of cellular proliferation (incorporated radioactivity)

28. Incorporation of ^{3}H-methyl thymidine is measured by β-scintillation counting as disintegrations per minute (DPM). Incorporation of ^{125}I-iododeoxyuridine is measured by ^{125}I-counting and also is expressed as DPM. The incorporation is expressed as DPM/mouse.

OBSERVATIONS

Clinical observations

29. Each mouse should be carefully observed at least once daily for any clinical signs, either of local irritation at the application site or of systemic toxicity. All observations are systematically recorded with records being maintained for each mouse. Monitoring plans should include criteria to promptly identify those mice exhibiting systemic toxicity, excessive irritation, or corrosion of skin for euthanasia.

Body weights

30. As stated in paragraph 26, individual animal body weights should be measured at the start of the test and at the scheduled kill.

CALCULATION OF RESULTS

31. Results for each treatment group are expressed as the mean SI. The SI is derived by dividing the mean DPM/mouse within each test substance group and the concurrent PC group by the mean DPM/mouse for the solvent/vehicle control group. The average SI for vehicle treated controls is then one.

32. The decision process regards a result as positive when SI $\geq$ 3. However, the strength of the dose-response, the statistical significance and the consistency of the solvent/vehicle and positive control responses may also be used when determining whether a borderline result is declared positive (4)(5)(6).

33. If it is necessary to clarify the results obtained, consideration should be given to various properties of the test substance, including whether it has a structural relationship to known skin sensitizers, whether it causes excessive local skin irritation in the mouse, and the nature of the dose-response seen. These and other considerations are discussed in detail elsewhere (7).

34. Collecting radioactivity data at the level of the individual mouse will enable a statistical analysis for presence and degree of dose response in the data. Any statistical assessment could include suitably adjusted comparisons of test groups (*e.g.* pair-wise dosed group versus concurrent vehicle control comparisons). Statistical analyses may include, for instance, linear regression or William's test to assess dose-response trends, and Dunnett's test for pair-wise comparisons. In choosing an appropriate method of statistical analysis, the investigator should maintain an awareness of possible inequalities of variances and other related problems that may necessitate a data transformation or a non-parametric statistical analysis. In any case the investigator may need to carry out SI calculations and statistical analyses with and without certain data points (sometimes called "outliers").

DATA AND REPORTING

Data

35. Data should be summarised in tabular form showing the individual animal DPM values, the group mean DPM/animal, its associated error term (*e.g.* SD, SEM), and the mean SI for each dose group compared against the concurrent vehicle control group.

Test report

36. The test report should contain the following information:

Test substance and control test substances:

- identification data (*e.g.* CAS number, if available; source; purity; known impurities; lot number);
- physical nature and physicochemical properties (*e.g.* volatility, stability, solubility);
- if mixture, composition and relative percentages of components.

Solvent/vehicle:

- identification data (purity; concentration, where appropriate; volume used);
- justification for choice of vehicle.

Test animals:

- source of CBA mice;
- microbiological status of the animals, when known;
- number and age of animals;
- source of animals, housing conditions, diet, etc.

Test conditions:

- details of test substance preparation and application;
- justification for dose selection (including results from pre-screen test, if conducted);
- vehicle and test substance concentrations used, and total amount of test substance applied;
- details of food and water quality (including diet type/source, water source);
- details of treatment and sampling schedules;
- methods for measurement of toxicity;
- criteria for considering studies as positive or negative;
- details of any protocol deviations and an explanation on how the deviation affects the study design and results.

Reliability check:

- a summary of results of latest reliability check, including information on test substance, concentration and vehicle used;
- concurrent and/or historical PC and concurrent negative control data for testing laboratory;
- if a concurrent PC was not included, the date and laboratory report for the most recent periodic PC and a report detailing the historical PC data for the laboratory justifying the basis for not conducting a concurrent PC.

Results:

- individual weights of mice at start of dosing and at scheduled kill; as well as mean and associated error term (*e.g.* SD, SEM) for each treatment group;
- time course of onset and signs of toxicity, including dermal irritation at site of administration, if any, for each animal;
- a table of individual mouse DPM values and SI values for each treatment group;
- mean and associated error term (*e.g.* SD, SEM) for DPM/mouse for each treatment group and the results of outlier analysis for each treatment group;
- calculated SI and an appropriate measure of variability that takes into account the inter-animal variability in both the test substance and control groups;
- dose response relationship;
- statistical analyses, where appropriate.

Discussion of results:

- a brief commentary on the results, the dose-response analysis, and statistical analyses, where appropriate, with a conclusion as to whether the test substance should be considered a skin sensitizer.

LITERATURE

(1) OECD (2002), Skin Sensitisation: Local Lymph Node Assay. OECD Guideline for the Testing of Chemicals No 429, Paris, France. Available at: [http://www.oecd.org/env/testguidelines].

(2) Kimber, I. and Basketter, D.A. (1992). The murine local lymph node assay; collaborative studies and new directions: A commentary. *Food Chem. Toxicol*., 30, 165-169.

(3) Kimber, I., Dearman, R.J., Scholes, E.W. and Basketter, D.A. (1994), The local lymph node assay: developments and applications. *Toxicol*., 93, 13-31.

(4) Kimber, I., Hilton, J., Dearman, R.J., Gerberick, G.F., Ryan, C.A., Basketter, D.A., Lea, L., House, R.V., Ladies, G.S., Loveless, S.E. and Hastings, K.L. (1998), Assessment of the skin sensitisation potential of topical medicaments using the local lymph node assay: An interlaboratory exercise. *J, Toxicol. Environ. Health*, 53, 563-79.

(5) Chamberlain, M. and Basketter, D.A. (1996), The local lymph node assay: status of validation. *Food Chem. Toxicol*., 34, 999-1002.

(6) Basketter, D.A., Gerberick, G.F., Kimber, I. and Loveless, S.E. (1996), The local lymph node assay: A viable alternative to currently accepted skin sensitisation tests. *Food Chem. Toxicol*., 34, 985-997.

(7) Basketter, D.A., Gerberick, G.F. and Kimber, I. (1998), Strategies for identifying false positive responses in predictive sensitisation tests. *Food Chem. Toxicol*., 36, 327-33.

(8) Van Och, F.M.M., Slob, W., De Jong, W.H., Vandebriel, R.J. and Van Loveren, H. (2000), A quantitative method for assessing the sensitising potency of low molecular weight chemicals using a local lymph node assay: employment of a regression method that includes determination of uncertainty margins. *Toxicol*., 146, 49-59.

(9) ICCVAM (2009), Recommended Performance Standards: Murine Local Lymph Node Assay. NIH Publication Number 09-7357. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna-ps/LLNAPerfStds.pdf].

(10) OECD (1992), Skin Sensitisation. OECD Guideline for Testing of Chemicals No 406, Paris, France. Available at: [http://www.oecd.org/env/testguidelines].

(11) OECD (2005), Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Environment, Health and Safety Monograph Series on Testing and Assessment No. 34. Available at: [http://www.oecd.org/env/testguidelines].

(12) Dearman, R.J., Hilton, J., Evans, P., Harvey, P., Basketter, D.A. and Kimber, I. (1998), Temporal stability of local lymph node assay responses to hexyl cinnamic aldehyde. *J. Appl. Toxicol*., 18, 281-284.

(13) Kimber, I., Dearman, R.J., Betts, C.J., Gerberick, G.F., Ryan, C.A., Kern, P.S., Patlewicz, G.Y. and Basketter, D.A. (2006), The local lymph node assay and skin sensitization: a cut-down screen to reduce animal requirements? *Contact Dermatitis*, 54, 181-185.

(14) ESAC (2007), Statement on the Reduced Local Lymph Node Assay (rLLNA). European Commission Directorate General, Joint Research Centre, Institute for Health and Consumer Protection, European Centre for the Validation of Alternative Methods. April 2007. Available at: [http://ecvam.jrc.it/ft_doc/ESAC26_statement_rLLNA_20070525-1.pdf].

(15) ICCVAM (2009), The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Test Method Evaluation Report. The Reduced Murine Local Lymph Node Assay: An Alternative Test Method Using Fewer Animals to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products. NIH Publication Number 09-6439. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at:

(16) ICCVAM (1999), The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds. The Results of an Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication No. 99-4494. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/llnarep.pdf].

(17) Kreiling, R., Hollnagel, H.M., Hareng, L., Eigler, L., Lee, M.S., Griem, P., Dreessen, B., Kleber, M., Albrecht, A., Garcia, C. and Wendel, A. (2008), Comparison of the skin sensitizing potential of unsaturated compounds as assessed by the murine local lymph node assay (LLNA) and the guinea pig maximization test (GPMT). *Food Chem. Toxicol*., 46, 1896-1904.

(18) Basketter, D., Ball, N., Cagen, S., Carrilo, J.C., Certa, H., Eigler, D., Garcia, C., Esch, H., Graham, C., Haux, C., Kreiling, R. and Mehling, A. (2009), Application of a weight of evidence approach to assessing discordant sensitisation datasets: implications for REACH. *Reg. Toxicol. Pharmacol*., 55, 90-96.

(19) ILAR (1996), Institute of Laboratory Animal Research Guide for the Care and Use of Laboratory Animals. 7th ed. Washington, DC: National Academies Press.

(20) McGarry, H.F. (2007), The murine local lymph node assay: regulatory and potency considerations under REACH. *Toxicol*., 238, 71-89.

(21) OECD (2002), Acute Dermal Irritation/Corrosion. OECD Guideline for Testing of Chemicals No.404, Paris, France: Available at: [http://www.oecd.org/env/testguidelines].

(22) Reeder, M.K., Broomhead, Y.L., DiDonato, L. and DeGeorge, G.L. (2007), Use of an enhanced local lymph node assay to correctly classify irritants and false positive substances. *Toxicologist*, 96 (S-1), 235.

(23) ICCVAM (2009), Non-radioactive Murine Local Lymph Node Assay: Flow Cytometry Test Method Protocol (LLNA: BrdU-FC) Revised Draft Background Review Document. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [*http://iccvam.niehs.nih.gov/methods/immunotox/fcLLNA/BRDcomplete.pdf*].

(24) Hayes, B.B., Gerber, P.C., Griffey, S.S. and Meade, B.J. (1998), Contact hypersensitivity to dicyclohexylcarbodiimide and diisopropylcarbodiimide in female B6C3F1 mice. *Drug. Chem. Toxicol*., 21, 195-206.

(25) Homey, B., von Schilling, C., Blumel, J., Schuppe, H.C., Ruzicka, T., Ahr, H.J., Lehmann, P. and Vohr, V.W. (1998), An integrated model for the differentiation of chemical-induced allergic and irritant skin reactions. *Toxicol. Appl. Pharmacol*., 153, 83-94.

(26) Woolhiser, M.R., Hayes, B.B. and Meade, B.J. (1998), A combined murine local lymph node and irritancy assay to predict sensitization and irritancy potential of chemicals. *Toxicol. Meth*., 8, 245-256.

(27) Hayes, B.B. and Meade, B.J. (1999), Contact sensitivity to selected acrylate compounds in B6C3F1 mice: relative potency, cross reactivity, and comparison of test methods. *Drug. Chem. Toxicol*., 22, 491-506.

(28) Ehling, G., Hecht, M., Heusener, A., Huesler, J., Gamer, A.O., van Loveren, H., Maurer, T., Riecke, K., Ullmann, L., Ulrich, P., Vandebriel, R. and Vohr, H.W. (2005), A European inter-laboratory validation of alternative endpoints of the murine local lymph node assay: first round. *Toxicol*., 212, 60-68.

(29) Vohr, H.W. and Ahr, H.J. (2005), The local lymph node assay being too sensitive? *Arch. Toxicol*., 79, 721-728.

(30) Patterson, R.M., Noga, E. and Germolec, D. (2007), Lack of evidence for contact sensitization by Pfiesteria extract. *Environ. Health Perspect*., 115, 1023-1028.

(31) OECD (1987), Acute Dermal Toxicity. OECD Guideline for Testing of Chemicals No 402, Paris, France. Available at: [http://www.oecd.org/env/testguidelines].

(32) ICCVAM (2009). Report on the ICCVAM-NICEATM/ECVAM/JaCVAM Scientific Workshop on Acute Chemical Safety Testing: Advancing In Vitro Approaches and Humane Endpoints for Systemic Toxicity Evaluations. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/methods/acutetox/Tox_workshop.htm].

ANNEX 1

PERFORMANCE STANDARDS FOR ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED LOCAL LYMPH NODE ASSAY TEST METHODS FOR SKIN SENSITISATION

INTRODUCTION

1. The purpose of Performance Standards (PS) is to communicate the basis by which new test methods, both proprietary (*i.e.* copyrighted, trademarked, registered) and non-proprietary can be determined to have sufficient accuracy and reliability for specific testing purposes. These PS, based on validated and accepted test methods, can be used to evaluate the reliability and accuracy of other analogous test methods (colloquially referred to as "me-too" tests) that are based on similar scientific principles and measure or predict the same biological or toxic effect (11).

2. Prior to adoption of modified test methods (*i.e.* proposed potential improvements to an approved test method), there should be an evaluation to determine the effect of the proposed changes on the test's performance and the extent to which such changes affect the information available for the other components of the validation process. Depending on the number and nature of the proposed changes, the generated data and supporting documentation for those changes, they should either be subjected to the same validation process as described for a new test, or, if appropriate, to a limited assessment of reliability and relevance using established PS (11).

3. Similar (me-too) or modified test methods proposed for use under this Test Guideline should be evaluated to determine their reliability and accuracy using chemicals representing the full range of the LLNA scores. To avoid unwarranted animal use, it is strongly recommended that model developers contact OECD before starting validation studies in accordance with the PS and guidance provided in this Test Guideline.

4. These PS are based on the ICCVAM/ECVAM/JaCVAM harmonised PS (9), for evaluating the validity of new or modified versions of the LLNA. The PS consists of essential test method components, recommended reference substances, and standards for accuracy and reliability that the proposed test method should meet or exceed.

I. Essential test method components

5. To ensure that a modified LLNA test method is functionally and mechanistically similar to the LLNA and measures the same biological effect, the following components should be included in the test method protocol:

1. The test substance should be applied topically to both ears of the mouse.
2. Lymphocyte proliferation should be measured in the lymph nodes draining from the site of test substance application.
3. Lymphocyte proliferation should be measured during the induction phase of skin sensitisation.

4. For test substances, the highest dose selected should be the maximum concentration that does not induce systemic toxicity and/or excessive local skin irritation in the mouse. For positive reference substances, the highest dose should be at least as high as the LLNA EC3 values of the corresponding reference test substances (see Table 1) without producing systemic toxicity and/or excessive local skin irritation in the mouse.
5. A concurrent vehicle control should be included in each study and, where appropriate, a concurrent PC should also be used.
6. A minimum of four animals per dose group is required.

If any of these criteria are not met, then these performance standards cannot be used for validation of the modified test method.

II. Minimum list of reference substances

6. The ICCVAM/ECVAM/JaCVAM harmonized PS (9) identified 18 minimum required reference substances and four optional reference substances (*i.e.* substances that produced either false positive or false negative results in the LLNA, when compared to human and guinea pig results (*i.e.* TG 406) (10), and therefore provide the opportunity to demonstrate equal to or better performance than the LLNA) that are included in the LLNA performance standards. The selection criteria for identifying these substances were:

- The list of reference substances represented the types of substances typically tested for skin sensitization potential and the range of responses that the LLNA is capable of measuring or predicting;
- The substances had well-defined chemical structures;
- LLNA data from guinea pig tests (*i.e.* TG 406) (10) and (where possible) data from humans were available for each substance; and
- The substances were readily available from a commercial source.

The recommended reference substances are listed in Table 1. Studies using the proposed reference substances should be evaluated in the vehicle with which they are listed in Table 1. In situations where a listed substance may not be available, other substances that meet the selection criteria mentioned may be used, with adequate justification.

556 **TABLE 1. RECOMMENDED REFERENCE SUBSTANCES FOR THE LLNA PERFORMANCE STANDARDS**

Number	Substance[1]	CAS No	Form	Veh[2]	EC3 %[3]	N[4]	0.5x - 2.0x EC3	Actual EC3 Range	LLNA vs. GP	LLNA vs. Human
1	5-Chloro-2-methyl-4-isothiazolin-3-one (CMI)/ 2-methyl-4-isothiazolin-3-one (MI)[5]	26172-55-4/ 2682-20-4	Liq	DMF	0.009	1	0.0045-0.018	NC	+/+	+/+
2	DNCB	97-00-7	Sol	AOO	0.049	15	0.025-0.099	0.02-0.094	+/+	+/+
3	4-Phenylenediamine	106-50-3	Sol	AOO	0.11	6	0.055-0.22	0.07-0.16	+/+	+/+
4	Cobalt chloride	7646-79-9	Sol	DMSO	0.6	2	0.3-1.2	0.4-0.8	+/+	+/+
5	Isoeugenol	97-54-1	Liq	AOO	1.5	47	0.77-3.1	0.5-3.3	+/+	+/+
6	2-Mercaptobenzothiazole	149-30-4	Sol	DMF	1.7	1	0.85-3.4	NC	+/+	+/+
7	Citral	5392-40-5	Liq	AOO	9.2	6	4.6-18.3	5.1-13	+/+	+/+
8	HCA	101-86-0	Liq	AOO	9.7	21	4.8-19.5	4.4-14.7	+/+	+/+
9	Eugenol	97-53-0	Liq	AOO	10.1	11	5.05-20.2	4.9-15	+/+	+/+
10	Phenyl benzoate	93-99-2	Sol	AOO	13.6	3	6.8-27.2	1.2-20	+/+	+/+
11	Cinnamic alcohol	104-54-1	Sol	AOO	21	1	10.5-42	NC	+/+	+/+
12	Imidazolidinyl urea	39236-46-9	Sol	DMF	24	1	12-48	NC	+/+	+/+
13	Methyl methacrylate	80-62-6	Liq	AOO	90	1	45-100	NC	+/+	+/+
14	Chlorobenzene	108-90-7	Liq	AOO	NA	1	NA	NA	-/-	-/*
15	Isopropanol	67-63-0	Liq	AOO	NA	1	NA	NA	-/-	-/+
16	Lactic acid	50-21-5	Liq	DMSO	NA	1	NA	NA	-/-	-/*
17	Methyl salicylate	119-36-8	Liq	AOO	NA	9	NA	NA	-/-	-/-
18	Salicylic acid	69-72-7	Sol	AOO	NA	1	NA	NA	-/-	-/-

Number	Substance[1]	CAS No	Form	Veh[2]	EC3 %[3]	N[4]	0.5x - 2.0x EC3	Actual EC3 Range	LLNA vs. GP	LLNA vs. Human
Optional Substances to Demonstrate Improved Performance Relative to the LLNA										
19	Sodium lauryl sulfate	151-21-3	Sol	DMF	8.1	5	4.05-16.2	1.5-17.1	+/-	+/-
20	Ethylene glycol dimethacrylate	97-90-5	Liq	MEK	28	1	14-56	NC	+/-	+/+
21	Xylene	1330-20-7	Liq	AOO	95.8	1	47.9-100	NC	+/**	+/-
22	Nickel chloride	7718-54-9	Sol	DMSO	NA	2	NA	NA	-/+	-/+

Abbreviations: AOO = acetone: olive oil (4:1); CAS No = Chemical Abstracts Service Number; DMF = *N*,*N*-dimethylformamide; DMSO = dimethyl sulfoxide; DNCB = 2,4-dinitrochlorobenzene; EC3 = estimated concentration needed to produce a stimulation index of 3; GP = guinea pig test result (*i.e.* TG 406) (10); HCA = hexyl cinnamic aldehyde; Liq = liquid; LLNA = murine local lymph node assay result (*i.e.* TG 429) (1); MEK = methyl ethyl ketone; NA = not applicable since stimulation index <3; NC = not calculated since data was obtained from a single study; Sol = solid; Veh = test vehicle.

[1] Test substances should be prepared daily unless stability data demonstrate the acceptability of storage.

[2] Because of the potential impact of different vehicles on the performance of the LLNA, the recommended vehicle for each reference substance should be used (20)(28).

[3] Mean value where more than one EC3 value was available.

[4] Number of LLNA studies from which data were obtained.

[5] Commercially available as Kathon CG (CAS No 55965-84-9), which is a 3:1 mixture of CMI and MI. The relative concentrations of each component range from 1.1% to 1.25% (CMI) and 0.3% to 0.45% (MI). The inactive components are magnesium salts (21.5% to 24%) and copper nitrate (0.15% to 0.17%), with the remaining formulation 74% to 77% water. Kathon CG is readily available through Sigma-Aldrich and Rohm and Haas (now Dow Chemical Corporation).

* = Presumed to be a non-sensitizer in humans based on the fact that no clinical patch test results were located, it is not included as a patch test kit allergen, and no case reports of human sensitisation were located.

** = GP data not available.

III. Defined reliability and accuracy standards

7. The accuracy of a modified LLNA test method should meet or exceed that of the LLNA PS when it is evaluated using the 18 minimum required reference substances. The new or modified test method should result in the correct classification based on a "yes/no" decision. However, the new or modified test method might not correctly classify all of the minimum required reference substances. If, for example, one of the weak sensitizers were misclassified, a rationale for the misclassification and appropriate additional data (*e.g.* test results that provide correct classifications for other substances with physical, chemical, and sensitizing properties similar to those of the misclassified reference substance) could be considered to demonstrate equivalent performance. Under such circumstances, the validation status of the new or modified LLNA test method would be evaluated on a case-by-case basis.

Intra-laboratory reproducibility

8. To determine intra-laboratory reproducibility, a new or modified LLNA test method should be assessed using a sensitizing substance that is well characterized in the LLNA. Therefore, the LLNA PS are based on the variability of results from repeated tests of hexyl cinnamic aldehyde (HCA). To assess intra-laboratory reliability, threshold estimated concentration (ECt) values for HCA should be derived on four separate occasions with at least one week between tests. Acceptable intra-laboratory reproducibility is indicated by a laboratory's ability to obtain, in each HCA test, ECt values between 5% and 20%, which represents the range of 0.5-2.0 times the mean EC3 specified for HCA (10%) in the LLNA (see Table 1).

Inter-laboratory reproducibility

9. Inter-laboratory reproducibility of a new or modified LLNA test method should be assessed using two sensitizing substances that are well characterized in the LLNA. The LLNA PS are based on the variability of results from tests of HCA and 2,4-dinitrochlorobenzene (DNCB) in different laboratories. ECt values should be derived independently from a single study conducted in at least three separate laboratories. To demonstrate acceptable inter-laboratory reproducibility, each laboratory should obtain ECt values of 5% to 20% for HCA and 0.025% to 0.1% for DNCB, which represents the range of 0.5-2.0 times the mean EC3 concentrations specified for HCA (10%) and DNCB (0.05%), respectively, in the LLNA (see Table 1).

ANNEX 2

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with "concordance" to mean the proportion of correct outcomes of a test method.

Benchmark test substance: A sensitizing or non-sensitizing substance used as a standard for comparison to a test substance. A benchmark substance should have the following properties; (i) consistent and reliable source(s); (ii) structural and functional similarity to the class of substances being tested; (iii) known physical/chemical characteristics; (iv) supporting data on known effects, and (v) known potency in the range of the desired response.

Estimated concentration threshold (ECt): Estimated concentration of a test substance needed to produce a stimulation index that is indicative of a positive response.

Estimated concentration three (EC3): Estimated concentration of a test substance needed to produce a stimulation index of three.

False negative: A test substance incorrectly identified as negative or non-active by a test method, when in fact it is positive or active.

False positive: A test substance incorrectly identified as positive or active by a test, when in fact it is negative or non-active.

Hazard: The potential for an adverse health or ecological effect. The adverse effect is manifested only if there is an exposure of sufficient level.

Inter-laboratory reproducibility: A measure of the extent to which different qualified laboratories, using the same protocol and testing the same test substances, can produce qualitatively and quantitatively similar results. Inter-laboratory reproducibility is determined during the prevalidation and validation processes, and indicates the extent to which a test can be successfully transferred between laboratories, also referred to as between-laboratory reproducibility.

Intra-laboratory reproducibility: A determination of the extent that qualified people within the same laboratory can successfully replicate results using a specific protocol at different times. Also referred to as within-laboratory reproducibility.

Me-too test: A colloquial expression for a test method that is functionally and mechanistically similar to a validated and accepted reference test method. Such a test method would be a candidate for catch-up validation. Interchangeably used with similar test method.

Outlier: An outlier is an observation that is markedly different from other values in a random sample from a population.

Performance standards (PS): Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is functionally and mechanistically similar. Included are; (i) essential test method components; (ii) a minimum list of Reference Chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (iii) the similar levels of accuracy

and reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals.

Proprietary test method: A test method for which manufacture and distribution is restricted by patents, copyrights, trademarks, etc.

Quality assurance: A management process by which adherence to laboratory testing standards, requirements, and record keeping procedures, and the accuracy of data transfer, are assessed by individuals who are independent from those performing the testing.

Reference chemicals: Chemicals selected for use in the validation process, for which responses in the *in vitro* or *in vivo* reference test system or the species of interest are already known. These chemicals should be representative of the classes of chemicals for which the test method is expected to be used, and should represent the full range of responses that may be expected from the chemicals for which it may be used, from strong, to weak, to negative. Different sets of reference chemicals may be required for the different stages of the validation process, and for different test methods and test uses.

Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method.

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility.

Skin sensitization: An immunological process that results when a susceptible individual is exposed topically to an inducing chemical allergen, which provokes a cutaneous immune response that can lead to the development of contact sensitization.

Stimulation Index (SI): A value calculated to assess the skin sensitization potential of a test substance that is the ratio of the proliferation in treated groups to that in the concurrent vehicle control group.

Test substance: Any material tested using this TG, whether it is a single compound or consists of multiple components (*e.g.* final products, formulations). When testing formulations, consideration should be given to the fact that certain regulatory authorities only require testing of the final product formulation. However, there may also be testing requirements for the active ingredient(s) of a product formulation.

Validated test method: A test method for which validation studies have been completed to determine the relevance (including accuracy) and reliability for a specific purpose. It is important to note that a validated test method may not have sufficient performance in terms of accuracy and reliability to be found acceptable for the proposed purpose.